



基于拓扑有效连通路径的有向网络 链路预测方法

李治成*, 吉立新, 刘树新, 李 星, 李劲松

(中国人民解放军战略支援部队信息工程大学 郑州 450001)

【摘要】链路预测旨在利用已有的网络拓扑信息来挖掘未知连边, 具有较高的应用价值。大部分已有的基于拓扑结构的方法, 关注节点对之间的路径数或者预测节点对的出入度, 未有效挖掘节点对之间的连边长度和连边上节点的影响力对相似性的影响。针对此问题, 该文提出了基于拓扑有效连通路径的链路预测方法, 并分析了不同路径长度在节点度、半局部中心性和 H-指数这 3 种不同衡量节点影响力指标下对节点相似性的贡献。通过 8 个真实网络仿真, 发现 H-指数能有效量化节点的局部影响力, 且在 3 种衡量指标下均具有较高的预测精度。

关键词 有向网络; 有效连通路径; 链路预测; 节点影响力

中图分类号 TP399 **文献标志码** A **doi:**10.12178/1001-0548.2020220

A Method of Link Prediction in Directed Network Based on Effective Connectivity Path

LI Zhi-cheng*, JI Li-xin, LIU Shu-xin, LI Xing, and LI Jin-son

(People's Liberation Army Strategic Support Force Information Engineering University Zhengzhou 450001)

Abstract Link prediction aims to mine unknown links based on observed topology information, which has high application value in many fields. At present, existing link prediction methods mainly focus on undirected network while the research of directed network is less. The prediction method based on structural information assumes that the more similar the nodes are, the more likely they are to be linked. Actually, the links between nodes are generated through paths, which cause similarity transfers between nodes. Most of the existing methods based on topology often focus on either the path between node pairs or the node degree, do not effectively mine the link length between node pairs and the local influence of nodes on the path. To solve this problem, this paper proposes a link prediction algorithm based on the effective connectivity path, and analyzes the contribution of different path length and node degree, semi-local centrality and H-index to node similarity. Compared with the existing eight prediction methods, the method proposed based on H-index effectively quantifies the local influence of nodes, and has a higher prediction accuracy under three indices.

Key words directed network; effective connectivity path; link prediction; node influence

随着网络科学的发展, 复杂网络成为了重要的研究内容^[1-5]。生活中的很多复杂系统都可以抽象为网络来表示, 不同类型的网络如交通网络、引文网络、蛋白质作用关系网络、社交网络等均是复杂网络的研究对象。而链路预测作为复杂网络的研究方向, 旨在利用已有的网络结构发现网络中缺失的连边、错误的连边和未来可能产生的连边^[6-8]。链路预测应用于诸多领域, 如在社交网络中用于好友推荐^[9], 生物网络中用于发现未知的生物结构^[10], 以及引文网络中发现科学家之间的合作关系^[11]等。

当前, 链路预测的研究已经取得较显著的成

果。基于拓扑结构的链路预测方法因简单、高效而备受关注^[12-14], 根据拓扑结构信息可将链路预测方法分为基于局部结构、基于半局部结构和基于全局结构 3 种方法^[15]。基于局部结构信息的预测方法假设如果节点越相似, 产生连接的可能性越大, 其中以共同邻居结构为出发点已有较多的研究方法, AA 指标对共同邻居的大度节点进行惩罚, CAR 方法考虑共同邻居节点之间相互交互^[15]。文献 [16] 对资源分配指标 RA 进行了改进, 提出扩展的资源分配指标。全局信息考虑网络的全局结构, 例如考虑全局路径的 Katz 方法^[17], 基于随机游走的节点偏好

收稿日期: 2020-05-10; 修回日期: 2020-10-29

基金项目: 国家自然科学基金(61803384)

作者简介: 李治成(1996-), 男, 主要从事复杂网络、链路预测、关键节点挖掘等方面的研究. E-mail: jtlizhicheng@163.com

性游走指标 (degree-biased random walk, DRW)^[18] 和有重启的随机游走指标 (random walk with restart, RWR)^[19] 等, 全局方法预测结果优于局部指标, 但因其时间复杂度高不适用于大型网络。文献 [20] 提出的局部路径方法 (local path, LP) 在共同邻居的基础上考虑了三阶邻居所起的作用, 在大部分网络中预测结果与 Katz 方法相近。目前已有大量的链路预测方法, 但现存的方法忽略了预测节点对之间的连边长度和连边上不同节点的影响力对相似性的贡献。节点对之间的相似性不仅与路径的长度有关, 还与路径上的节点的影响力有关, 节点影响力挖掘方法不同, 连边传递的有效性也将不同。此外, 还需探讨不同挖掘方法下节点影响力对相似性的贡献。

因此, 本文分析了网络在不同路径长度和节点影响力下对节点相似性的贡献, 提出基于拓扑有效连通路径的链路预测方法。而节点在网络中影响力的挖掘方法较多, 为减少时间复杂度, 本文主要从节点的出入度、半局部中心性和 H-指数 3 种中心性衡量指标来量化节点的局部影响力, 并使用相应的方法来量化连边的传递能力。通过在多种实际网络中进行实验, 验证了节点之间的相似性不仅与路径长度有关, 还与节点影响力有关。

1 相关工作介绍

有向网络中, 节点相似性的指标较多, 常用的几种相似性指标对比如下:

1) Bifan^[21]: 该方法基于网络势能理论提出, 通过筛选出网络中出现频率最高的模体结构并应用于连边的预测, 表达式为:

$$\mathbf{S} = \mathbf{A}\mathbf{A}'\mathbf{A} \quad (1)$$

式中, \mathbf{S} 表示相似性矩阵; \mathbf{A} 表示邻接矩阵; \mathbf{A}' 表示 \mathbf{A} 的转置矩阵。

2) 共同邻居 (directed common neighbors, DCN)^[22]: 利用节点对 x 、 y 之间的共同邻居来衡量节点的相似性, 该方法假设如果节点对之间的共同邻居越多, 节点对之间存在连边的可能性越大, 表示为:

$$s(x, y) = |\Gamma_{\text{out}}(x) \cap \Gamma_{\text{in}}(y)| \quad (2)$$

式中, $\Gamma_{\text{out}}(x)$ 表示节点 x 的出边邻居节点集合; $\Gamma_{\text{in}}(y)$ 表示节点 y 的入边邻居节点集合。

3) DAA^[22]: 在 DCN 的基础上, 考虑共同邻居对节点相似性的贡献。计算公式为:

$$s(x, y) = \sum_{z \in \Gamma_{\text{out}}(x) \cap \Gamma_{\text{in}}(y)} \frac{1}{\log k_z^{\text{out}}} \quad (3)$$

式中, k_z^{out} 表示节点 z 的出度数。

4) 资源分配指标 (directed resource allocation, DRA)^[22]: 该方法基于资源分配的假设, 资源在传输过程中, 平均分配给节点的每一条出边, 邻居节点最终获得的资源量与该节点的出边成反比, 表示为:

$$s(x, y) = \sum_{z \in \Gamma_{\text{out}}(x) \cap \Gamma_{\text{in}}(y)} \frac{1}{k_z^{\text{out}}} \quad (4)$$

5) 偏好性连接指标 (directed preference attachment, DPA)^[22]: 该方法假设两个节点之间产生连边的可能性与各自的出入度相关。具体表示为:

$$s(x, y) = k_x^{\text{out}} \bullet k_y^{\text{in}} \quad (5)$$

式中, k_y^{in} 表示节点 y 的入度数。

6) 局部相似性指标 (LP)^[20]: 在二阶有向路径的基础上考虑三阶有向路径所起的作用, 并对三阶路径进行了惩罚, 可表示为:

$$\mathbf{S} = \mathbf{A}^2 + \alpha \mathbf{A}^3 \quad (6)$$

式中, α 为惩罚系数, 一般取值为 0.01。

7) Katz^[17]: 从网络的全局属性出发, 计算连接节点 x 能到达 y 的所有路径, 表示为:

$$s(x, y) = \sum_{l=1}^{\infty} a^l \bullet |\text{paths}_{x,y}^{(l)}| \quad (7)$$

式中, $\text{paths}_{x,y}^{(l)}$ 为节点 x 和节点 y 之间长度为 l 的路径数目; a 为调节路径的权重。

8) 平均通勤时间 (ACT)^[23]: 游走的粒子从节点 x 游走至节点 y 所需的平均步数, 表示为:

$$s(x, y) = \frac{1}{L_{xx}^+ + L_{yy}^+ - 2L_{xy}^+} \quad (8)$$

式中, L_{xy}^+ 为对应的拉普拉斯矩阵的伪逆 L^+ 中对应的元素值。

9) 局部随机游走方法 (local random walk, LRW)^[24]: 假设粒子在 t 时刻从 x 出发, $\pi_{xy}(t)$ 表示粒子在 $t+1$ 时刻正好到节点 y 的概率, 则演化方程表示为:

$$\pi_x(t+1) = \mathbf{P}^T \pi_x(t) \quad 0 \leq t \quad (9)$$

式中, \mathbf{P} 表示转移概率矩阵。

假设各个节点初始资源分布为 \mathbf{q}_x , 基于 t 步随机游走的相似性表示为:

$$s(x, y) = \mathbf{q}_x \bullet \pi_{xy}(t) + \mathbf{q}_y \bullet \pi_{yx}(t) \quad (10)$$

10) 有叠加效应的局部随机游走 (superposed random walk, SRW)^[24]: 该方法在 LRW 基础上将 t 步及其以前的结果相加, 表示为:

$$s(x, y) = \sum_{l=1}^l s_{xy}^{LRW}(l) \quad (11)$$

11) propflow^[25]: 该方法是一种有监督的随机游走方法, 计算从源节点 x 开始以一定步长或更少的步长到达节点 y 的概率。

2 基于拓扑有效连通路程的有向网络链路预测方法

节点之间通过路径产生联系, 网络中往往存在多条路径, 路径的拓扑连通度是影响节点之间的相似性传递的主要原因。链路预测中, 路径的拓扑连通度是指信息通过路径传输, 最终能传递到目标节点的能力大小, 路径连通度越大, 节点之间越相似。以社交网络中信息传递为例, 一个用户有较多好友使得他拥有较高的影响力, 该用户可以将信息通过好友传递给其他用户。在这个过程中, 信息将通过好友之间构成的关系网络完成传递, 好友影响力越大, 能传递信息的能力越强, 但不能保证信息能传递到某个具体的用户, 因为路径的连通度不同, 导致用户获取到的信息也不同。然而, 目前基于路径的链路预测方法缺乏对路径长度与节点影响力的分析, 本文将从拓扑有效连通路程分析网络中节点之间产生连接的可能性。首先分析影响节点相似性的重要因素, 在此基础上分析了路径连通度对节点相似性的贡献, 并提出了预测方法。

2.1 拓扑连通度的分析量化

传统的链路预测方法研究中, 认为端点的影响力有助于连边的形成, 且影响力大的节点之间更倾向于连接^[26]。而网络中端点间实际通过路径建立联系, 并通过建立的关系来传递影响力, 从而发生相似性传递。而网络中相似节点对之间往往存在多条路径, 路径结构的差异将影响相似性的传递。图 1 给出了拓扑结构与信息传输路径的示意图。信息从多条路径传递到 y , 结构的差异将影响路径连通度大小。对于相同长度路径, $x \rightarrow a \rightarrow y$ 和 $x \rightarrow b \rightarrow y$, 传递到 y 的信息与中继节点有关, 中继节点的入度不同, 连边的连通度也不同。对于不同长度的路径 $x \rightarrow a \rightarrow y$ 和 $x \rightarrow b \rightarrow e \rightarrow f \rightarrow y$, 分别通过 2 跳和 4 跳到达, 节点信息在传递过程中往往随着长度的增加而减少, 不如短路径传递的信息有效。而最终 y 获取到的有效信息是多条路径的叠加。因此, 需要从拓扑结构对信息传输过程影响的角度, 分析刻画拓扑路径的连通度。

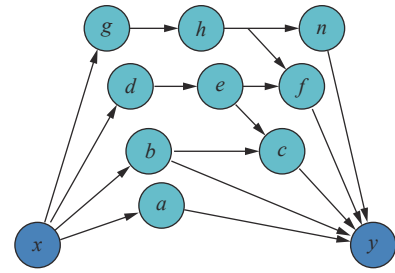


图 1 网络中信息传输过程分析示意图

为了探讨拓扑路径连通度对相似性传递的影响, 图 2 构造了 4 个拓扑结构图进行分析。一般情况下, 节点之间的路径数是影响路径连通度的重要因素, 节点之间路径越多, 路径连通度越大, 相比于图 2b, 图 2a 中的 x 和 y 之间具有较高连通度。此外, 路径长度是影响相似性传递的重要因素, 当信息从 x 开始传递时, 图 2c 中传递信息的路径有 $x \rightarrow e \rightarrow f \rightarrow y$ 和 $x \rightarrow f \rightarrow y$, 传递过程需经过节点 f , 但 $x \rightarrow f \rightarrow y$ 比 $x \rightarrow e \rightarrow f \rightarrow y$ 高效, 因此 $x \rightarrow e \rightarrow f \rightarrow y$ 所起的作用较少。而在图 2d 的 $x \rightarrow v \rightarrow u \rightarrow \dots \rightarrow n \rightarrow y$ 路径中, 传递信息的过程中会随着长度的增加而耗散, 最终传递到 y 的信息微乎其微。因此, 拓扑路径的连通度需有效区分长短路径在信息传递过程中的作用。

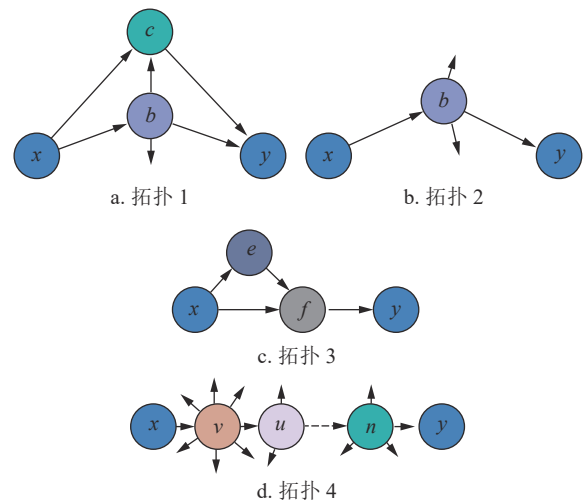


图 2 不同路径长度下相似性传递过程分析

其次, 节点间路径的长短虽然重要, 但路径中节点的差异性也决定了路径能传递的信息多少, 还需分析路径连通度与中继节点的关系。节点影响力描述了节点在网络中的重要程度, 且节点影响力能够决定节点在路径中捕获资源和信息的能力^[27], 路径中传递资源往往偏向于影响力大的节点。文献 [28] 证实了在电力网络中, 节点的度与承受的电力负荷成正比, 现实生活中“富者愈富”的现象也说明影响力大的节点在网络中占有优势^[29]。在有向网络中,

相连节点之间的路径连通度与中继节点的出入边的影响力有关。在图3所示的网络结构中,如果仅以节点出入度作为判断节点影响力的大小,则节点 z_2 具有较高影响力,节点 x 传递资源时偏向于 z_2 ,但由于影响力大的中继节点能传递的连边较多,最终能传递到相邻节点 y 的资源较小,即中继节点影响力低且路径短的预测节点之间的连边连通度较大。

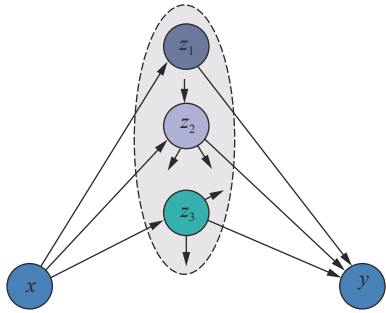


图3 中继节点与路径连通度的关系分析

综上所述,影响节点相似性的重要原因是路径长度和中继节点的影响力,中继节点影响连边的传递能力,进而影响路径连通度。基于此,首先对节点和邻居节点之间传递的容量进行量化,突出与入边影响力大的节点之间的连边传递容量,并从出边和入边两个角度量化节点的传输容量。因此,有向网络 $G(V,E)$ 节点 x 传递到邻居节点 y 的连边连通度可量化为:

$$w_{xy} = \frac{AT_x \bullet AC_y}{\sum_{k \in \Gamma_{out}(x)} AT_x \bullet AC_k} \quad (12)$$

式中, AT_x 表示节点 x 的出边影响力; AC_y 表示节点 y 的入边影响力。式(12)仅定义了通用的量化方法来刻画连边的连通度,具体量化将在后文给出。其次,对于网络中未连接的节点,应突出影响力小的节点在连通路径的作用,并区分不同长度路径所起的作用,故节点 x 和 y 之间的单条路径连通度可以量化为:

$$C(x,y) = \frac{w_{xc}}{\sum_{k \in \Gamma_{out}(x)} w_{xk}} \frac{w_{cu}}{\sum_{j \in \Gamma_{out}(c)} w_{cj}} \cdots \frac{w_{vy}}{\sum_{u \in \Gamma_{out}(v)} w_{vu}} \quad (13)$$

可以看出,任意一条路径信息传输的有效性可以量化为路径长度和路径中节点影响力对信息传递能力的影响。然而,有向网络中影响力的挖掘方法较多,不同网络中,中心性衡量存在差异,因此,本文还分析了在不同节点度中心性指标下对路径连通度的贡献。目前已有的中心性方法并不能完全挖

掘节点在网络中的影响力,为减少时间复杂度,主要利用局部结构方法定义节点的重要程度,选取基于节点的度中心性、节点的半局部中心性和H-指数3种中心性指标量化节点的入边和出边影响力,并给出了相应方法的路径连通度的表达式,具体方法定义如下。

基于节点度的中心性指标:仅考虑节点出入度,节点的影响力表达式为:

$$d_{out}(v) = \sum_{u \in V} a_{vu} \quad (14)$$

$$d_{in}(v) = \sum_{u \in V} a_{uv} \quad (15)$$

式中, $d_{out}(v)$ 表示节点 v 的出度; $d_{in}(v)$ 表示节点的入度。网络中邻接矩阵表示为 $A = \{a_{vu}\}$,当矩阵中 a_{vu} 为1时表示节点 v 和 u 之间有连边,为0表示节点对之间不存在连边。对于未连接的节点 x 和 y ,单条路径 $x \rightarrow e \cdots v \rightarrow y$ 的连通度表示为:

$$C(x,y) = \frac{d_{out}(x) \bullet d_{in}(e)}{\sum_{k \in \Gamma_{out}(x)} d_{out}(x) \bullet d_{in}(k)} \cdots \frac{d_{out}(v) \bullet d_{in}(y)}{\sum_{u \in \Gamma_{out}(v)} d_{out}(v) \bullet d_{in}(u)} \quad (16)$$

基于节点半局部中心性的中心性指标:在节点度的基础上考虑了邻居节点的重要性,出边和入边重要性的表达式分别为:

$$LC_{out}(v) = \sum_{u \in \Gamma_{out}(v)} \sum_{w \in \Gamma_{in}(u)} |\Gamma_{in}(w)| \quad (17)$$

$$LC_{in}(v) = \sum_{u \in \Gamma_{in}(v)} \sum_{w \in \Gamma_{out}(u)} |\Gamma_{out}(w)| \quad (18)$$

基于半局部中心性指标中,未连接的节点 x 和 y ,单条路径 $x \rightarrow e \cdots v \rightarrow y$ 连通度表示为:

$$C(x,y) = \frac{LC_{out}(x) \bullet LC_{in}(e)}{\sum_{k \in \Gamma_{out}(x)} LC_{out}(x) \bullet LC_{in}(k)} \cdots \frac{LC_{out}(v) \bullet LC_{in}(y)}{\sum_{u \in \Gamma_{out}(v)} LC_{out}(v) \bullet LC_{in}(u)} \quad (19)$$

H-指数(H-index)^[30]:H-指数反应了论文的质量,用于评估作者发表论文与论文质量之间的关系。本文计算H-指数时仅考虑了节点的一阶邻居节点,表示为:

$$h_{in}^1(v) = H \{d_{out}(j_1), d_{out}(j_2), \cdots, d_{out}(j_{d_{in}(v)})\} \quad (20)$$

$$h_{out}^1(v) = H \{d_{in}(s_1), d_{in}(s_2), \cdots, d_{in}(s_{d_{out}(v)})\} \quad (21)$$

式中, $j_1, j_2, \cdots, j_{d_{in}(v)}$ 表示指向节点 v 的邻居节点;

$s_1, s_2, \dots, s_{d_{out}(v)}$ 表示节点 v 指向的邻居节点。相应地, 基于 H-指数量化路径连通度的方法中, 任意两个未连接的节点 x 、 y , 节点 x 通过若干节点到达目标节点 y 的单条路径连通度计算公式为:

$$C(x, y) = \frac{h_{out}^1(x) \cdot h_{in}^1(e) \dots h_{out}^1(v) \cdot h_{in}^1(y)}{\sum_{k \in \Gamma_{out}(x)} h_{out}^1(x) \cdot h_{in}^1(k) \dots \sum_{u \in \Gamma_{out}(v)} h_{out}^1(v) \cdot h_{in}^1(u)} \quad (22)$$

2.2 预测方法提出

为了从路径有效性对信息传输影响的角度刻画任意两点间的相似性, 将基于路径传输影响因素分析节点间信息传输能力。“三度影响力”认为人与人之间相互影响, 且影响力的传递是在三度以内, 超出三度节点, 影响力将会逐渐消失^[31], 当源节点和目标节点之间经过三跳或者更多跳路径时, 相似性传递也将随着路径长度的增加而无法有效的传递到目标节点。“六度分割理论”认为世界上任何两个陌生人之间的距离不超过 6 个人, 也就是说, 每个人最多通过 6 个人就可以认识世界上任何一个陌生人^[32]。兼顾“三度影响力”和“六度分割理论”, 将两度节点和六度节点之间的路径称为拓扑有效连通路径, 并从信息传递的角度定义了相关方法。

定义 1 基于拓扑有效连通路径的有向网络链路预测方法

对于给定的有向网络 $G(V, E)$, 任意两个未连接节点 x 和 y , 两个节点之间拓扑有效连通路径可用于量化节点之间相似性, 其相似性与节点影响力和路径长度有关, 具体表示为:

$$s(x, y) = \sum_{path_{x,y}^l} \frac{w_{xc}}{\sum_{k \in \Gamma_{out}(x)} w_{xk}} \frac{w_{cu}}{\sum_{j \in \Gamma_{out}(c)} w_{uj}} \dots \frac{w_{vy}}{\sum_{u \in \Gamma_{out}(v)} w_{vu}} \quad 1 < path \leq 6 \quad (23)$$

式中, $path_{x,y}^l$ 表示节点 x 和 y 之间长度为 l 的路径, 该方法计算网络中任意未连接的节点 x 和 y 路径连通度的和。其本质是一种随机游走的方法, 需满足节点在游走过程中到达信息已传递过的节点时停止游走。式 (23) 为基于拓扑有效连通路径的通用公式。为便于比较不同节点度中心性方法之间的差异, 对节点度的中心性指标、节点半局部中心性指标和 H-指数量化路径中节点影响力的方法分别进行定义。

基于节点度中心性的有效连通路径方法 (connectivity path based on degree, EPD) 分别从节点的

出度和入度两个角度量化路径中的节点出边和入边的影响力, 对于任意未连接的节点 x 和 y , 节点间相似性的表达式为:

$$s(x, y) = \sum_{path_{x,y}^l} \frac{d_{out}(x) \cdot d_{in}(e) \dots d_{out}(v) \cdot d_{in}(y)}{\sum_{k \in \Gamma_{out}(x)} d_{out}(x) \cdot d_{in}(k) \dots \sum_{u \in \Gamma_{out}(v)} d_{out}(v) \cdot d_{in}(u)} \quad 1 < path \leq 6 \quad (24)$$

基于节点半局部中心性的有效连通路径方法 (connectivity path based on semi-local centrality, EPLC) 在 EPD 的基础上, 考虑了网络中的二阶邻居节点数目。在该方法中, 未连接节点 x 和 y 之间的连边有效连通表示为:

$$s(x, y) = \sum_{path_{x,y}^l} \frac{LC_{out}(x) \cdot LC_{in}(e) \dots LC_{out}(v) \cdot LC_{in}(y)}{\sum_{k \in \Gamma_{out}(x)} LC_{out}(x) \cdot LC_{in}(k) \dots \sum_{u \in \Gamma_{out}(v)} LC_{out}(v) \cdot LC_{in}(u)} \quad 1 < path \leq 6 \quad (25)$$

相比于节点度, H-指数在计算节点相似性时, 考虑了邻居的重要性。网络中任意未连接的节点 x 和 y , 基于 H-指数的有效连通路径方法 (connectivity path based on H-index, EPH) 计算节点相似性的表达式为:

$$s(x, y) = \sum_{path_{x,y}^l} \frac{h_{out}^1(x) \cdot h_{in}^1(e) \dots h_{out}^1(v) \cdot h_{in}^1(y)}{\sum_{k \in \Gamma_{out}(x)} h_{out}^1(x) \cdot h_{in}^1(k) \dots \sum_{u \in \Gamma_{out}(v)} h_{out}^1(v) \cdot h_{in}^1(u)} \quad 1 < path \leq 6 \quad (26)$$

3 衡量指标和数据集介绍

3.1 衡量指标

对于给定的链路预测方法, 通过计算节点对之间的相似性得分来衡量未连接的节点对之间的相似性, 本文通过 AUC (area under receiver operating characteristic curve)、precision、排序分 (ranking score, RS) 来衡量预测精确度, 将有向网络 $G(V, E)$ 的连边 E 分为训练集 E^T 和测试集 E^P , 满足 $E^P \cup E^T = E$, 且 $E^P \cap E^T = \emptyset$ 。

AUC 为随机从测试集 E^P 中随机选取一条边的分数值比未连接边的分数值大的概率^[33]。如果选取的测试边的分数值有 n' 条大于未连接边的分数值, 则加 n' 分, 若分数值相等的有 n'' 条, 则加 $0.5n''$ 分, AUC 的计算公式表达为:

$$\text{AUC} = \frac{n' + 0.5n''}{n} \quad (27)$$

精确度 (precision) 衡量指标为在前 L 条边中预测准确的比例^[34]。如果在前 L 条边中有 m 条在测试集出现, 则精确度表示为:

$$\text{precision} = \frac{m}{L} \quad (28)$$

一般情况下, L 的取值为 100。为对比不同规模网络中预测性能, 文中 L 取值为连边数的 10%。

RS 从最终的边的排序位置来衡量预测精确度, 假设 $H = U - E^T$ 为位置边的集合, 测试边 r_e 表示测试边在排序中的排名, 可得测试边的排序分为:

$$\text{RS}_e = \frac{r_e}{|H|} \quad (29)$$

系统的排序分为:

$$\text{RS} = \frac{1}{|EP|} \sum_{e \in EP} \text{RS}_e = \frac{1}{|EP|} \sum_{e \in EP} \frac{r_e}{|H|} \quad (30)$$

3.2 网络数据介绍

本文选取了不同类型的有向网络用于预测验证, 数据集介绍如下:

1) FWEW^[35]: 生活在 Everglades Graminoids 的湿季生物构成的食物链网络, 节点表示物种, 连边表示物种之间的捕食关系;

2) High-school (HIG)^[36]: 描述美国伊利诺伊州某高中男生间建立的朋友关系网络。数据集中每个用户被问谁是朋友, $A \rightarrow B$ 表示用户 A 认为 B 是 A 的朋友;

3) Residence_hall (RES)^[36]: 澳大利亚国立大学朋友关系网络, 与 HIG 类似;

4) CollegeMessage (CM)^[36]: 一个由加州大学欧文分校构成的邮件网络构成的社交网络;

5) Political_blogs (PB)^[37]: 美国政治博客网络, 节点表示网页, 连边表示网页之间的连边关系;

6) Scimet^[38]: 引用“科学计量学”的论文引文网络, 节点表示论文, 连边表示论文引用关系;

7) SmaGri2^[38]: 一个由 Small & Griffith and Descendants 相关的论文引用网络, 节点表示论文, 连边表示引用关系;

8) Air traffic control (AIR)^[36]: 由美国国家联邦航空局飞行数据中心构建的航空网络, 节点表示机场或者服务中心, 连边表示飞行路线。

表 1 使用网络数据参数说明

| No. | Network | Type | V | E | MID | MOD | $\langle k \rangle$ | $\langle d \rangle$ | C |
|-----|---------|----------------|-------|--------|-------|-----|---------------------|---------------------|-------|
| 1 | FWEW | Ecological | 69 | 916 | 63 | 44 | 13.3 | 2.8 | 0.552 |
| 2 | HIG | Social | 70 | 366 | 18 | 12 | 10.5 | 3.5 | 0.404 |
| 3 | RES | Social | 128 | 2 137 | 110 | 63 | 16.7 | 0.1 | 0.335 |
| 4 | CM | Social | 1 899 | 59 835 | 1 091 | 558 | 31.5 | 6.2 | 0.109 |
| 5 | PB | Hyperlink | 1 222 | 19 021 | 337 | 256 | 4.1 | 3.6 | 0.320 |
| 6 | Scimet | Citation | 2 678 | 10 381 | 121 | 104 | 3.9 | 6.1 | 0.174 |
| 7 | SmaGri2 | Citation | 1 024 | 4 919 | 89 | 232 | 4.8 | 4.9 | 0.302 |
| 8 | AIR | Transportation | 1 226 | 2 615 | 24 | 28 | 4.3 | 6.1 | 0.064 |

上述的网络具体参数特征如表 1 所示, 包括网络类型 Type、节点数 $|V|$ 、边数 $|E|$ 、最大入度 MID、最大出度 MOD、平均度 $\langle k \rangle$ 、平均距离 $\langle d \rangle$ 、聚类系数 C 。实验中, 设置训练集和测试集所占的比例为 9:1, 每个实验结果均为 50 次实验结果的均值。

4 实验分析

为了验证路径的有效传播在链路预测方法中的作用, 在 8 个真实网络上进行实验分析。实验首先分析路径中节点影响力与拓扑路径有效连通的关系, 并对不同路径长度下的有效连通度进行对比。

4.1 中心度方法之间对比分析

图 4 为 EPD、EPLC、EPH 在 8 个真实网络中

的 AUC 性能曲线, 本文取节点之间最长的路径为 10, 以验证节点之间的有效路径连通度是否在 6 步以内。首先对比不同的网络中各个节点中心度量方法之间的差异度, 根据曲线变化可以看出 EPLC 的预测结果波动性较大, 在有的网络中表现较好, 有的网络中表现较差, 在 EPD 和 EPH 表现相对较好, 且 EPH 优于 EPD。因此, 相比于节点度中心、节点的半局部中心的中心衡量指标, H-指数更适用于量化节点的影响力。信息在传递过程中, 并不是所有量化方法都能准确衡量出节点的影响力, H-指数考虑邻居节点个数的同时, 还考虑了邻居节点的质量, 能有效计算连边传递的相似性, 因此该

方法表现较好且稳定。EPD 仅根据节点的出入度量化节点的影响力, 并不能充分挖掘节点的重要程度。EPLC 在 EPD 的基础上考虑了局部邻居节点的数目, 近似地计算了节点在全局网络中的相对作用大小, 但应用于链路预测的时候, 预测精度较低, 且稳定性较差。

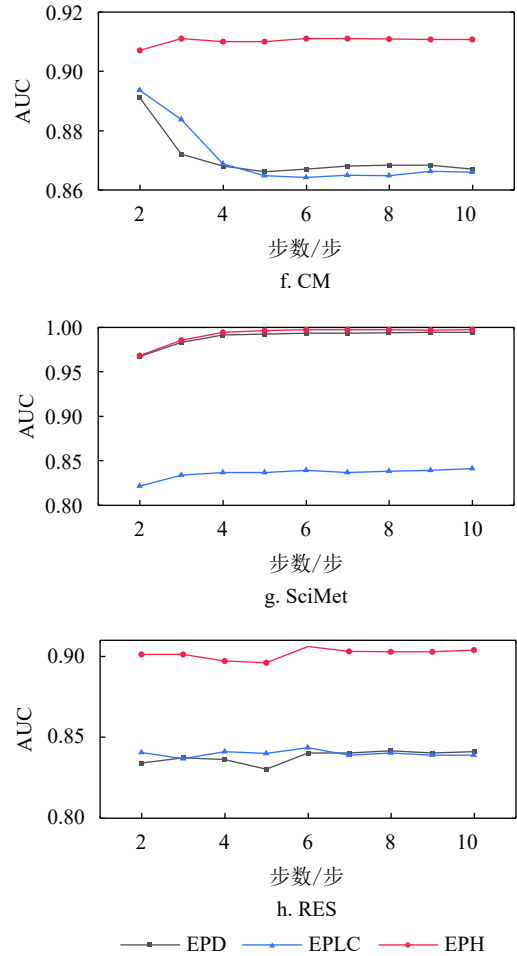
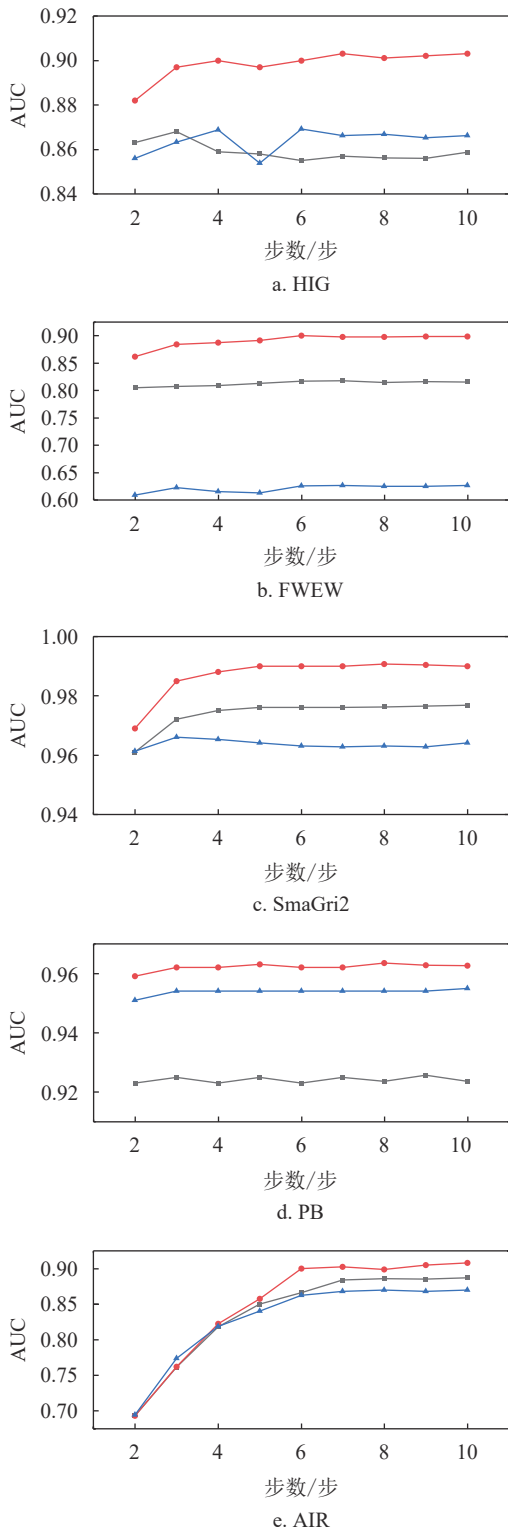


图 4 不同中心度方法之间 AUC 对比分析

不同游走步数下, AUC 表现各不相同, 说明不同路径长度下对节点的相似性贡献不同。当游走长度为 2 时, 计算的是粒子从源节点出发两步到达目标节点的相似度, 从预测结果中可以看出, 大部分网络中的 AUC 预测精度并不是很高。在 CM 网络中, 由于网络中节点具有较高的平均度, 和节点之间平均距离较大, 当网络中具有较大度节点时并不能促进相似节点产生, 当路径长度为 2~4 时, EPD 方法没有准确对连边的连通度进行衡量, 在计算节点之间的相似性得分的时候, 增加了干扰信息, 造成 EPD 方法在网络中随着路径的增长预测精度下降, 但当路径长度大于 4 时候, 路径的连通度变得微乎其微, 因此, 路径长度为 4 之后预测精度逐渐趋于稳定。而在 Air 航空网络中平均距离较大, 当一个机场有多条路径可达一个机场时, 说明该机场可能是一个大型国际机场, 目标节点越重要, 与其产生连接的可能性越大, 所以 AUC 结果会增加。可以看出, 在所有网络中, 3 种节点量化节点重要性的方法均随着路径长度的增加而改变, 在大部分网络中, 当路径长度达到 6 时基本趋于稳

定。因此，节点之间的拓扑路径有效连通路程在 2 度和 6 度节点之间。

4.2 不同预测方法对比分析

基于不同中心度量化节点影响力的实验结果可以看出，当游走的步长为 6 时 AUC 值趋于稳定，为了便于与其他链路预测方法进行对比分析，同时兼顾六度分割理论，EPD、EPLC、EPH 游走的步长 l 值均为 6，使用 AUC、precision、排序分 3 种衡量指标进行评价。

4.2.1 AUC 实验对比分析

表 2 给出了 AUC 的对比分析，Katz.01 表示调节参数 α 的取值为 0.01，LRW_3 中 3 表示游走的步数为 3。通过对比分析可以得出，基于 H-指数量化的方法表现较好，在 Scimet、SmaGri2、AIR 网络中 EPD、EPLC 均比 EPH 差，说明在一些网络中影响力的有效传播与连边的量化方式相关，并不是所

有的量化方法都能有效量化连边影响力。4 种局部指标 DCN、DAA、DRA、DPA 考虑的网络结构简单，在各种类型的网络中表现差异较大。而 LP 在共同邻居 (DCN) 的基础上考虑了三阶路径所起的作用，有较大的改进。全局方法 Katz 因考虑全局网络的拓扑结构信息均表现最好。而随机游走方法局部随机游走和有叠加效应的随机游走方法在不同的网络中表现差距较大，在规模相对较大的网络 Scimet 中表现较好，小规模网络中预测的 AUC 值几乎接近于 0，而 ACT 方法在所有预测方法中是最差的，说明相似性在传递过程中并不一定能传递到目标节点，且该方法时间复杂度较高。Bifan 在食物链方法中预测精度较高，超过了考虑全局信息的 Katz 方法，而此网络结构并不适用于其他类型的网络，说明食物链网络在演化过程中产生较多的 Bifan 结构，能有效提升食物链网络中的预测精度。

表 2 AUC 结果分析

| Dataset | FWEW | HIG | RES | CM | PB | Scimet | SmaGri2 | AIR |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Bifan | 0.924 | 0.839 | 0.865 | 0.904 | 0.960 | 0.940 | 0.960 | 0.641 |
| DCN | 0.784 | 0.859 | 0.886 | 0.752 | 0.927 | 0.895 | 0.908 | 0.609 |
| DAA | 0.785 | 0.808 | 0.896 | 0.736 | 0.863 | 0.891 | 0.876 | 0.595 |
| DRA | 0.774 | 0.862 | 0.891 | 0.755 | 0.928 | 0.895 | 0.909 | 0.609 |
| DPA | 0.901 | 0.660 | 0.670 | 0.910 | 0.944 | 0.957 | 0.966 | 0.856 |
| LP | 0.830 | 0.889 | 0.899 | 0.901 | 0.958 | 0.965 | 0.971 | 0.696 |
| Katz.01 | 0.834 | 0.893 | 0.899 | 0.905 | 0.960 | 0.995 | 0.990 | 0.900 |
| ACT | 0.669 | 0.681 | 0.687 | 0.722 | 0.753 | 0.652 | 0.706 | 0.762 |
| LRW_3 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 | 0.416 | 0.000 | 0.000 |
| SRW_3 | 0.019 | 0.000 | 0.000 | 0.000 | 0.000 | 0.418 | 0.000 | 0.000 |
| propflow | 0.81 | 0.861 | 0.886 | 0.875 | 0.938 | 0.995 | 0.979 | 0.865 |
| EPD | 0.817 | 0.850 | 0.840 | 0.867 | 0.923 | 0.993 | 0.976 | 0.866 |
| EPLC | 0.626 | 0.866 | 0.843 | 0.864 | 0.954 | 0.977 | 0.963 | 0.862 |
| EPH | 0.900 | 0.900 | 0.906 | 0.911 | 0.962 | 0.997 | 0.990 | 0.900 |

纵向对比分析发现在航空网络 AIR 数据集上，Katz.01 时预测精度较高，而 LP 方法预测结果与其相差接近 0.2，说明在该网络中，路径越长，航班之间直飞的可能性越大。而改进后的 EPD、EPH 预测结果与 Katz.01 相差不大，考虑路径的异构性确实能促进节点对之间的相似。在 DCN 基础上改进的 DAA 指标，在 RES 和 AIR 网络中预测精度提高较大，大度节点在这两个网络中抑制节点对之间的相似连接，而在其余网络中预测结果均无改进。对于大部分网络，PA 方法结果具有较高的预测精度，很好的解释了“富者愈富”的现象，节点更倾向于和大度节点相连接。而本文方法针对不同的网络结构，路径的有效连通度并不相同，在各种连边量化的方法中 EPH 方法具有较高的预测精度，所有网络均具有较好的表现，能较好地量化节

点对之间的连边的连通度，也进一步说明有效的邻居节点确实能提高节点之间的相似性。

4.2.2 precision 结果分析

表 3 给出了 precision 的结果，可以看出，预测结果与 AUC 结果差不多，食物链网络中依旧 bifan 结构最高，4 种局部结构中，PA 在食物链网络中表现较好，LP 和 Katz 方法的预测精度较为接近；3 种随机游走方法 ACT、LRW_3、SRW_3 预测准确度较低，但在社交网络中表现较好，而在大型网络中预测精度几乎为 0；在小规模网络的社交网络 RES、HIG 预测结果较好，可能随机游走的指标在大规模网络中更适合 AUC。在 CM 网络中，相比 Katz0.01 预测精度提升了 1%，且时间复杂度远小于 Katz。在 CN 基础上改进的 AA、RA 指标，在引文网络 Scimet、Smagri2 和航

空网络中精度较差, 引文网络和航空网络节点的度与重要程度成正比, 重要节点并不会抑制网络连边的 L 形成。考虑了有效路径连通度之后, EPH 的 precision 结果在除了生物网络之外的所有

网络中表现均较好, 说明节点的影响力不仅与节点自身有关, 还与邻居节点的重要性相关, 在量化邻居节点重要性的基础上, EPH 方法能有效计算路径的连通度。

表 3 precision 结果分析

| Dataset | FWEW | HIG | RES | CM | PB | Scimet | SmaGri2 | AIR |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| bifan | 0.479 | 0.127 | 0.127 | 0.012 | 0.192 | 0.114 | 0.128 | 0.038 |
| DCN | 0.145 | 0.161 | 0.196 | 0.009 | 0.189 | 0.032 | 0.134 | 0.031 |
| DAA | 0.178 | 0.154 | 0.207 | 0.009 | 0.193 | 0.060 | 0.121 | 0.015 |
| DRA | 0.262 | 0.154 | 0.204 | 0.012 | 0.067 | 0.025 | 0.078 | 0.016 |
| DPA | 0.233 | 0.022 | 0.037 | 0.014 | 0.190 | 0.110 | 0.069 | 0.007 |
| LP | 0.146 | 0.127 | 0.180 | 0.008 | 0.188 | 0.115 | 0.128 | 0.030 |
| Katz.01 | 0.145 | 0.127 | 0.180 | 0.006 | 0.190 | 0.114 | 0.127 | 0.030 |
| ACT | 0.055 | 0.005 | 0.034 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| LRW_3 | 0.000 | 0.175 | 0.009 | 0.001 | 0.002 | 0.000 | 0.000 | 0.003 |
| SRW_3 | 0.000 | 0.180 | 0.009 | 0.001 | 0.002 | 0.000 | 0.000 | 0.003 |
| propflow | 0.321 | 0.139 | 0.153 | 0.001 | 0.035 | 0.034 | 0.032 | 0.004 |
| EPD | 0.324 | 0.082 | 0.092 | 0.001 | 0.055 | 0.037 | 0.029 | 0.004 |
| EPLC | 0.049 | 0.005 | 0.001 | 0.001 | 0.011 | 0.000 | 0.021 | 0.000 |
| EPH | 0.171 | 0.175 | 0.183 | 0.015 | 0.195 | 0.117 | 0.135 | 0.040 |

4.2.3 排序分结果分析

表 4 为排序分的实验结果, 与 AUC、precision 预测结果不同, 该衡量指标的预测结果越小, 精度越高, 即预测结果越排在前, 证明方法的可行性越好。从表 4 的预测结果中可以看出, EPH 方法表现的结果依旧较好, 局部指标 CN、AA、RA 预测结果较为接近, 因为这 3 种方法网络结构相同。在不同的度中心衡量方法中, EPLC 的预测结果最差, 说明该方法并不能有效量化路径连通度, 考虑

的网络结构不适合应用于该链路预测中。而随机游走方法 LRW_3、SRW_3 偏差较大, 在 SmaGri2、AIR 网络, 排序结果均错误, 可能与取的游走方向和步数有关。相比于局部方法, PA 排序结果较好, 对大部分网络能准确地将连边出现的概率进行排序。在所改进的方法中, EPLC 在 EPD 的基础上考虑了节点在局部结构中的相对重要性, 排序结果较差。综合对比, 在局部网络结构中, 使用 H-指数量化路径中节点影响力的方法最为有效。

表 4 排序分结果分析

| Dataset | FWEW | HIG | RES | CM | PB | Scimet | SmaGri2 | AIR |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| bifan | 0.094 | 0.178 | 0.140 | 0.087 | 0.039 | 0.081 | 0.048 | 0.224 |
| DCN | 0.290 | 0.139 | 0.108 | 0.085 | 0.055 | 0.132 | 0.129 | 0.121 |
| DAA | 0.273 | 0.138 | 0.104 | 0.230 | 0.054 | 0.131 | 0.146 | 0.235 |
| DRA | 0.274 | 0.137 | 0.105 | 0.211 | 0.055 | 0.043 | 0.129 | 0.231 |
| DPA | 0.140 | 0.389 | 0.347 | 0.210 | 0.057 | 0.042 | 0.057 | 0.138 |
| LP | 0.206 | 0.113 | 0.099 | 0.078 | 0.039 | 0.004 | 0.039 | 0.198 |
| Katz.01 | 0.199 | 0.115 | 0.100 | 0.095 | 0.039 | 0.004 | 0.012 | 0.093 |
| ACT | 0.406 | 0.381 | 0.337 | 0.086 | 0.248 | 1.000 | 0.293 | 0.236 |
| LRW_3 | 0.844 | 0.105 | 0.331 | 0.279 | 0.204 | 1.000 | 1.000 | 0.328 |
| SRW_3 | 0.848 | 0.089 | 0.329 | 0.526 | 0.204 | 1.000 | 1.000 | 0.327 |
| propflow | 0.231 | 0.123 | 0.123 | 0.122 | 0.063 | 0.006 | 0.021 | 0.121 |
| EPD | 0.223 | 0.162 | 0.171 | 0.131 | 0.077 | 0.007 | 0.024 | 0.121 |
| EPLC | 0.289 | 0.587 | 0.667 | 0.171 | 0.237 | 0.024 | 0.037 | 0.173 |
| EPH | 0.130 | 0.109 | 0.095 | 0.076 | 0.035 | 0.003 | 0.009 | 0.113 |

5 结束语

近年来, 基于网络拓扑结构相似性的链路预测

方法备受学者关注, 已提出了较多方法。大部分方法将节点的度构建为节点的影响力, 影响力越大,

与之产生连接的可能性越大,但连边的建立本质上与路径的有效性连通有关,路径上中继节点的影响力将影响连边的建立。本文将节点影响力应用于有向网络链路预测中,通过多种方法量化路径的连通能力。在多个实际网络中进行实验,结果表明,对于相同的游走步数,用基于H-指数更适合表示节点的影响力,能有效量化路径的有效连通度,而当有多条路径可达目标节点时,3阶以上路径对节点之间的相似性传递贡献较小。本文所提出的方法说明路径的有效连通度不仅与路径长度有关,还与路径上的节点影响力有关。此外,该预测方法能应用于大型网络,能有效地揭示连边的形成机理和网络的演化机制。

参 考 文 献

- [1] HAGHANI S, KEYVANPOUR M R. A systemic analysis of link prediction in social network[J]. *Artificial Intelligence Review*, 2019, 52(3): 1961-1995.
- [2] 刘树新, 季新生, 刘彩霞, 等. 一种信息传播促进网络增长的网络演化模型[J]. *物理学报*, 2014, 2014(15): 1-11.
LIU Shu-xin, JI Xin-sheng, LIU Cai-xia, et al. A complex network evolution model for network growth promoted by information transmission[J]. *Acta Physica Sinica*, 2014, 2014(15): 1-11.
- [3] 王凯, 李星, 兰巨龙, 等. 一种基于资源传输路径拓扑有效性的链路预测方法[J]. *电子与信息学报*, 2020, 42(3): 653-660.
WANG Kai, LI Xing, LAN Ju-long, et al. A new link prediction method for complex networks based on topological effectiveness of resource transmission paths[J]. *Journal of Electronics and Information Technology*, 2020, 42(3): 653-660.
- [4] LI Jin-song, PENG Jian-hua, LIU Shu-xin, et al. Link prediction in directed networks utilizing the role of reciprocal links[J]. *IEEE Access*, 2020, 8: 28668-28680.
- [5] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: A survey[J]. *IEEE Internet of Things Journal*, 2017, 5(1): 450-465.
- [6] BÜTÜN E, KAYA M. A pattern based supervised link prediction in directed complex networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 525: 1136-1145.
- [7] ZHANG Peng, QIU Dan, ZENG An, et al. A comprehensive comparison of network similarities for link prediction and spurious link elimination[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 500: 97-105.
- [8] KUMAR A, SINGH S S, SINGH K, et al. Link prediction techniques, applications, and performance: A survey[J]. *Physica A: Statistical Mechanics and Its Applications*, 2020, 553: 124289.
- [9] KUNDU S, PAL S K. Double bounded rough set, tension measure, and social link prediction[J]. *IEEE Transactions on Computational Social Systems*, 2018, 5(3): 841-853.
- [10] CUI Ying, CAI Meng, DAI Ying, et al. A hybrid network-based method for the detection of disease-related genes[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 492: 389-394.
- [11] LIU Han-wen, KOU Huai-zhen, YAN Cao, et al. Link prediction in paper citation network to construct paper correlation graph[J]. *EURASIP Journal on Wireless Communications and Networking*, 2019, 2019(1): 1-12.
- [12] LIU Shu-xin, JI Xin-sheng, LIU Cai-xia, et al. Similarity indices based on link weight assignment for link prediction of unweighted complex networks[J]. *International Journal of Modern Physics B*, 2017, 31(2): 1650254.
- [13] WU Yi-teng, YU Hong-tao, HUANG Rui-yang, et al. A fusion link prediction method based on limit theorem[J]. *Applied Sciences*, 2018, 8(1): 32.
- [14] AGHABOZORGI F, KHAYYAMBASHI M R. A new similarity measure for link prediction based on local structures in social networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 501: 12-23.
- [15] MARTINEZ V, BERZAL F, CUBERO J C. A survey of link prediction in complex networks[J]. *ACM Computing Surveys (CSUR)*, 2016, 49(4): 1-33.
- [16] LIU Shu-xin, JI Xin-sheng, LIU Cai-xia, et al. Extended resource allocation index for link prediction of complex network[J]. *Physica A: Statistical Mechanics and its Applications*, 2017, 479: 174-183.
- [17] KATZ L. A new status index derived from sociometric analysis[J]. *Psychometrika*, 1953, 18(1): 39-43.
- [18] ZHANG Yun-yi, SHI Zhan, FENG Dan, et al. Degree-Biased random walk for large-scale network embedding[J]. *Future Generation Computer Systems*, 2019, 100: 198-209.
- [19] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[J]. *Computer Networks and ISDN Systems*, 1998, 30(1-7): 107-117.
- [20] ZHOU Tao, LÜ Lin-yuan, ZHANG Yi-cheng. Predicting missing links via local information[J]. *The European Physical Journal B*, 2009, 71(4): 623-630.
- [21] ZHANG Qian-ming, LÜ Lin-yuan, WANG Wen-qiang, et al. Potential theory for directed networks[J]. *PloS One*, 2013, 8(2): e55437.
- [22] ZHANG Xue, ZHAO Cheng-li, WANG Xiao-jie, et al. Identifying missing and spurious interactions in directed networks[J]. *International Journal of Distributed Sensor Networks*, 2015, 11(9): 507386.
- [23] KLEIN D J, RANDIC M. Resistance distance[J]. *Journal of Mathematical Chemistry*, 1993, 12(1): 81-95.
- [24] LIU Wei-ping, LÜ Lin-yuan. Link prediction based on local random walk[J]. *EPL (Europhysics Letters)*, 2010, 89(5): 58007.
- [25] LICHTENWALTER R N, LUSSIER J T, CHAWLA N V. New perspectives and methods in link prediction[C]// *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM, 2010: 243-252.
- [26] YANG Yu-jie, ZHANG Jian-hua, ZHU Xu-zhen, et al.

- Link prediction via significant influence[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 492: 1523-1530.
- [27] 李兰茜. 基于复杂网络结构的链路预测技术研究[D]. 北京: 北京邮电大学, 2019.
LI Lan-qian. Link prediction based on structure of complex network[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [28] WANG Xing-yuan, CAO Jian-ye, QIN Xiao-meng. Study of robustness in functionally identical coupled networks against cascading failures[J]. *PloS One*, 2016, 11(8): e0160545.
- [29] DEWHURST D R, DANFORTH C M, DODDS P S. Continuum rich-get-richer processes: Mean field analysis with an application to firm size[J]. *Physical Review E*, 2018, 97(6): 062317.
- [30] ZHAI Li, YAN Xiang-bin, ZHANG Guo-jing. The bi-directional h-index and B-core decomposition in directed networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 531: 121715.
- [31] CHRISTAKIS N A, FOWLER J H. Social contagion theory: Examining dynamic social networks and human behavior[J]. *Statistics in Medicine*, 2013, 32(4): 556-577.
- [32] GUARE J. Six degrees of separation: A play[M]. [S.l.]: Vintage, 1990.
- [33] ZENG Guo-ping, ZENG E. On the three-way equivalence of AUC in credit scoring with tied scores[J]. *Communications in Statistics-Theory and Methods*, 2019, 48(7): 1635-1650.
- [34] WU Zhi-hao, LIN You-fang, ZHAO Yi-ji, et al. Improving local clustering based top-L link prediction methods via asymmetric link clustering information[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 492: 1859-1874.
- [35] ULANOWICZ R E, DEANGELIS D L. Network analysis of trophic dynamics in south florida ecosystems[J]. *US Geological Survey Program on the South Florida Ecosystem*, 2005, 114: 45-48.
- [36] KUNEGIS J. KONECT network dataset[DB/OL]. [2020-01-12]. <http://konect.uni-koblenz.de/networks/>.
- [37] ADAMIC L A, GLANCE N. The political blogosphere and the 2004 US election: Divided they blog[C]// *Proceedings of the 3rd International Workshop on Link Discovery*. [S.l.]: [s.n.], 2005: 36-43.
- [38] BATAGELJ V, MRVAR A. Pajek —Analysis and visualization of large networks[M]//BATAGELJ V, MRVAR A. *Graph Drawing Software*. Berlin, Heidelberg: Springer, 2004: 77-103.

编辑 税 红